U.PORTO

MESTRADO INTEGRADO
MEDICINA

Addressing Voice Gender and Voice Feminization Training

Ivo Rúben Moreira Fernandes

M

JUNHO 2024

Addressing Voice Gender and Voice Feminization Training

Mestrado Integrado em Medicina

Instituto de Ciências Biomédicas Abel Salazar – Universidade do Porto (ICBAS-UP)

Estudante: Ivo Rúben Moreira Fernandes

Aluno de 6.º ano do Mestrado Integrado em Medicina (n.º mecanográfico: 201807721).

Endereço de correio eletrónico: Ivo.RMF@gmail.com

Ivo Rúben Moreira Fernandes

(Ivo Rúben Moreira Fernandes)

Orientador: Prof. Doutora Mariline Santos

Doutorada em Ciências Médicas pelo Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto.

Professora Associada Convidada do Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto.

Assistente Hospitalar em Otorrinolaringologia da Unidade Local de Saúde Santo António.

Endereço de correio eletrónico: marilinesantos2910@gmail.com

Mariline Santos

(Mariline Santos)

Co-orientadora: Prof. Doutora Marta Sofia Pinto Martins

Doutorada em Psicologia pela Faculdade de Psicologia e de Ciências da Educação, Universidade do Porto.

Professora Auxiliar Convidada do Iscte – Instituto Universitário de Lisboa (Iscte-IUL).

Investigadora Integrada do Centro de Investigação e Intervenção Social do Iscte-IUL.

Endereço de correio eletrónico: marta.sofia.martins@iscte-iul.pt

Marta Sofia Pinto Martins

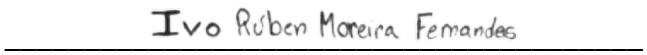
(Marta Sofia Pinto Martins)

02 de junho de 2024

DECLARAÇÃO DE INTEGRIDADE

Eu, Ivo Rúben Moreira Fernandes, n.º 201807721, aluno do Mestrado Integrado em Medicina, declaro ter atuado segundo os princípios sobre integridade académica da Universidade do Porto na elaboração da presente dissertação.

Nesse sentido, confirmo que na totalidade da elaboração deste trabalho não incorri na prática de plágio, de falsificação ou manipulação de resultados e com o cuidado necessário para o respeito aos grupos vulneráveis nele incluídos.



(Ivo Rúben Moreira Fernandes)
Universidade do Porto, 02 de junho de 2024

DEDICATÓRIA

To my mentor, Charles Robert Armstrong, III, thank you from the bottom of my heart for first showing me, years ago, that voice does not exist in an ethereal realm, free from the bounds of science. Anything about voice that I did not learn through you, I learned thanks to you. It's the brilliant minds of people like you, who choose to go against the grain, that allow for paradigm shifts like the one you're trying to replicate in the voice field. Who would've thought that speaking, singing, voice acting, and transgender voice could all fit under the same umbrella? Who would've thought art could join Biology and Medicine in such a monumental way? Thank you for always pushing me to be better, I will do the same.

To Scinguistics, for all the knowledge we built and shared together, for all the late nights talking about voice, singing, and getting frustrated with classical singing pedagogy.

À minha grande amiga Margarida Ione, por ser o meu grande apoio a navegar o mundo da academia e a voz da razão em tantas dimensões da minha vida.

Aos meus amigos, por todas as gargalhadas, todas as lágrimas e todas as lembranças que vão persistir até na memória de uma pessoa muito esquecida. Dedico esta dissertação também a vocês por me deixarem continuar a crescer convosco.

À minha família académica e à Comissão Organizadora da Oporto Biomedical Summit. Não consigo descrever em palavras o quão importantes vocês foram e continuam a ser no meu crescimento.

À Tuna Académica de Biomédicas, que me ensinou tanto e me permite levar comigo memórias únicas; por ser o grande local onde me posso expressar criativamente.

Ao ICBAS, por todas as oportunidades que me deu desde o meu primeiro ano. Foi no ICBAS que aprendi que muitas vezes as portas estão abertas.

A todos os professores que desde criança me transmitiram uma paixão enorme pelas ciências e humanidades, por serem um exemplo e modelo a seguir quando não tinha outro. Levo um bocado de cada um de vocês dentro de mim.

No fundo, pretendo dedicar este trabalho a todos os que se cruzaram por mim ou de alguma forma contribuíram para este trabalho. Cada encontro, cada palavra, cada gesto moldaram a pessoa que sou hoje e o trabalho que aqui apresento.

AGRADECIMENTOS

À minha orientadora, Professora Doutora Mariline Santos, obrigado por estar disposta a acolher um estudante de Medicina desconhecido, apenas por estar interessado em voz, desde a primeira cirurgia que me deixou assistir, há anos atrás, e por sempre me valorizar e incentivar, mesmo sendo eu apenas estudante. Obrigado pela confiança que depositou em mim no desenvolvimento de uma investigação tão complexa. Obrigado por começar a consulta de voz transgénero neste hospital e aceitar investigar comigo.

À minha coorientadora, Professora Doutora Marta Martins, pelo seu fascínio por esta área, por aceitar o desafio de começar uma investigação numa área tão diferente da sua (e que afinal acabou por ser muito semelhante); por todas as questões científicas que me coloca e que parecem sempre simplificar tudo na minha cabeça; pelo empréstimo do seu material de gravação durante tanto tempo; por me apresentar e explicar a estatística Bayesiana, que moldou esta investigação; pelas reuniões intermináveis a propor mil possíveis variáveis confundidoras para todas as questões científicas que nos surgem; por todas as ideias de projetos futuros; e por todos os "é uma questão empírica".

To Emilia, Lílian, and all my colleagues at Scinguistics, for sharing your knowledge and resources on voice acoustics with me. I never imagined how many full sets of 24 hours I'd end up spending learning Physics, all so I could learn to decipher the seemingly one-hundred different ways to extract each acoustic parameter, and better understand how the voice works, so I could simplify it in this Introduction and problematize it in the remaining text.

To Doctor Johan Sundberg, for his insight on vocal tract length estimation, and all the researchers who gladly shared manuscripts of their research with me, and for all our helpful correspondence.

A todos os meus amigos e família, que me ajudaram lendo a minha dissertação, participando na experiência de heteroperceção e até mesmo partilhando a experiência e recrutando participantes – antes sequer de eu o ter pedido, permitindo obter tantas respostas. Aos que me ouviram falar apaixonadamente sobre este projeto e sempre me perguntavam como estava.

Ao serviço de ORL e à equipa de audiometria por me acolher com tanta simpatia e permitir a utilização dos seus espaços e equipamento.

Ao Doutor Luís Meireles, Professora Dra. Isabel Palma e Professor Dr. Eurico Castro Alves por permitirem o desenvolvimento deste trabalho incorporado nos seus serviço, unidade e departamento, respetivamente.

RESUMO

Contexto: A voz é fundamental na identidade de género e no bem-estar de indivíduos transgénero. A terapia hormonal de afirmação de género tem-se revelado eficaz na masculinização da voz (em homens transgénero), mas não na feminização (em mulheres transgénero). Programas de treino vocal têm sido propostos como uma abordagem complementar para abordar esta questão.

Objetivos. Este estudo visou (1) investigar as características vocais e as experiências de indivíduos transgénero que frequentaram a consulta de voz transgénero na Unidade de Sexologia e Género da Unidade Local de Saúde Santo António, (2) determinar os preditores acústicos da heteroperceção do género da voz e (3) avaliar os efeitos do programa de treino de feminização vocal implementado na consulta.

Métodos. A amostra incluiu mulheres transgénero, homens transgénero, mulheres cisgénero e homens cisgénero, os quais completaram avaliações vocais, questionários sobre a sua perceção da voz (autoperceção) e qualidade de vida, e gravaram amostras de voz para análise acústica e para a avaliação heteropercetiva. Parte da amostra de mulheres transgénero (*n* = 7) foi selecionada por um terapeuta da fala independente para participar num programa de treino de feminização vocal. Para análise da heteroperceção do género da voz, ouvintes externos avaliaram estímulos de voz gravados.

Resultados. Mulheres e homens transgénero relataram uma qualidade de vida relacionada com a voz significativamente inferior à dos indivíduos cisgénero. A análise acústica revelou diferenças nos parâmetros vocais entre os grupos. O H1 residual corrigido para frequências de formantes e larguras de banda estava negativamente correlacionado com a frequência fundamental. As variáveis acústicas e sociodemográficas explicaram mais de 80% da variabilidade na heteroperceção numa análise de regressão. Para estímulos de vogais, a heteroperceção do género da voz foi determinada principalmente pela frequência fundamental, e para estímulos de frases isoladas e frases em texto, pela frequência fundamental e comprimento implícito do trato vocal. O programa de treino não mostrou efeito significativo na qualidade de vida relacionada com a voz, medida através do Índice de Desvantagem Vocal e *Trans Woman Voice Questionnaire*. Em relação à acústica e heteroperceção da voz, as mulheres transgénero mostraram uma voz mais semelhante à feminina na heteroperceção das vogais e uma frequência fundamental mais aguda em estímulos de frases em texto. No entanto, ambos os resultados permaneceram diferentes dos das mulheres cisgénero. Não foram encontradas diferenças no H1 residual*.

Conclusão. As características e experiências vocais diferem entre indivíduos transgénero e cisgénero. A heteroperceção do género da voz depende de diversos parâmetros acústicos. Portanto, é fundamental clarificar a relevância dos parâmetros acústicos isolados e a sua interação na definição do género da voz, e depois focar-se no desenvolvimento de programas de modificação do género da voz que visem diretamente esses parâmetros. Isto é essencial para atender às necessidades das mulheres transgénero, assim como dos homens transgénero que não respondem adequadamente à terapia hormonal de afirmação de género masculiniza.

Palavras-chave: género; pessoas transgénero; acústica; treino de feminização vocal; Português Europeu.

ABSTRACT

Background. Voice is key in gender identity and well-being of transgender individuals. Gender-affirming hormone therapy has proven to be effective in masculinizing the voice (in transgender men), but not in feminizing it (in transgender women). Voice training programs have been proposed as a complementary approach to address this issue.

Aims. This study aimed (1) to investigate the voice characteristics and experiences of transgender individuals attending the transgender voice consult at *Unidade de Sexologia e Género* of *Unidade Local de Saúde Santo António*, (to) to determine acoustic predictors of voice gender heteroperception, and (3) to assess the effects of the voice feminization training program implemented as part of the consult.

Methods. The sample included transgender women, transgender men, cisgender women, and cisgender men. Participants completed voice assessments, questionnaires on voice perception and quality of life, and recorded voice samples for acoustic analysis. A part of the TW sample (*n* = 7) was selected by an independent speech-language pathologist to undergo a voice feminization training program. Voice gender was assessed via recorded voice stimuli by external listeners.

Results. Transgender women and men reported significantly lower voice-related quality of life than the cisgender individuals. Acoustic analysis revealed differences in voice parameters between groups. Residual H1 corrected for formant frequencies and bandwidths was negatively correlated to fundamental frequency. Acoustic and sociodemographic variables explained more than 80% of the variability in heteroperception in a regression analysis. For vowel stimuli, voice gender heteroperception was mainly determined by f_o, and for single-phrase and phrase-in-text stimuli, by fo and implied vocal tract length. The training program showed no significant effect in voice-related quality of life measured via VHI and TWVQ. Regarding acoustics and heteroperception of voice, TW showed a more woman-like voice as perceived in the heteroperception of vowels, and a higher f_o in phrase-in-text stimuli. However, both results remained different from that of CW. No changes in residual H1* were found.

Conclusion. Voice characteristics and experiences differ between transgender and cisgender individuals. Voice gender heteroperception depends on diverse acoustic parameters. Thus, it is critical to clarify the relevance of isolated acoustics parameters and their interplay on defining voice gender, and then focus on developing voice gender modification programs that directly target those parameters. This is mandatory to meet the needs of transgender women, as well as of transgender men who do not respond adequately to masculinizing gender-affirming hormone therapy.

Keywords: gender; transgender persons; acoustics; voice feminization training; European Portuguese.

ACRONYMS

CM.	Cisgender men.
CW.	Cisgender women.
DVC.	Daily voice congruency.
f_o.	Fundamental frequency.
F3.	Frequency of formant 3.
F4.	Frequency of formant 4.
GAHT.	Gender-affirming hormone therapy.
IPA.	International Phonetic Alphabet.
iVTL.	Implied vocal tract length.
MIDI	Musical Instrument Digital Interface.
TM.	Transgender men.
TW.	Transgender women.
TWVQ.	Trans Woman Voice Questionnaire.
USEG-ULSSA.	*Unidade de Sexologia e Género* at *Unidade Local de Saúde Santo António*.
VHI.	Voice Handicap Index.

Index

TABLE LIST

FIGURE LIST

INTRODUCTION

Voice is an acoustic signal that is pivotal to the perception of gender identity, and critical to social presentation and affirmation. For transgender men and women, achieving a voice that aligns well with their gender identity is a primary concern that highly impacts their well-being and social integration (Hancock, 2017; Van Borsel G. De Cuypere, 2000). Despite its relevance, voice modification has been poorly explored within the gender-affirming process and transgender healthcare. Gender affirming process and transgender healthcare are multidisciplinary constructs and including voice within these frameworks will improve the comprehensiveness of the provided care.

Historical context. Voice modification in transgender individuals has been an incidental consequence of the physical alterations resulting from the gender-affirming hormone therapy (GAHT). This approach has proven effective in masculinizing the voice (Ziegler et al., 2018), but not for feminizing it (Bultynck et al., 2017). In addition, recent research suggests that masculinizing GAHT may not be effective for all transgender men — some experience an increase in voice misgendering when their voice is perceived by others (voice gender heteroperception), and/or an increase in dissatisfaction with their own voice (Hodges-Simeon et al., 2021; Ziegler et al., 2018). This is relevant because an atypical gender presentation puts transgender individuals at risk for external and self-inflicted violence (Barboza et al., 2016; Silva et al., 2022).

In respect to voice feminization, diverse approaches were tested. These approaches focused primarily on pitch, because women usually have higher pitched voices than men (Simpson, 2009), and addressed this acoustic parameter using surgical procedures and/or training programs (Kim, 2020). Recent evidence suggests, however, that a paradigm shift is mandatory, and that understanding voice apparatus and related acoustics is essential to appropriate voice modification procedures.

Theoretical basis of voice production. Voice production results from three modular and interconnected components: the Power, the Source, and the Filter. The Power comes from the lungs, diaphragm, and accessory respiratory muscles pushing air through the glottis, which is the space between the true vocal folds. The Source refers mainly to the true vocal folds, which move in a controlled, wave-like manner**.** This movement, that matches glottis opening and closing at a rate known as the fundamental frequency (f_o), produces regions of rarefaction (closed glottis) and compression (open glottis) of the outflowing air, i.e., forming a sound wave. The perceived pitch almost always matches the f_o. For example, a note pitched A4 has a frequency of 440 Hz, i.e., the vocal folds made an opening-closing cycle — a glottal cycle — at a rate of 440 times per second. In addition to the fundamental frequency (i.e., the first harmonic), a mathematically infinite number of harmonics is also produced. Each harmonic is an integer multiple of the fundamental frequency. The amplitude of each harmonic decreases across the harmonic series. Regarding the glottal cycle, when the vocal folds are in an open state for longer, they spend less time in collision, and that leads to a greater loss of amplitude across the harmonic series (Titze, 2009). The gradient defined by the harmonic frequencies is known as the spectral tilt. A low open quotient results in a higher spectral tilt (less negative), and a higher open quotient results in a lower spectral tilt (more negative).

As harmonics pass through the vocal tract, i.e., the Filter, they are differentially amplified by the resonance characteristics of the vocal tract geometry and tissues and group together into regions with different peaks, known as formants. Vowels are defined by the frequencies of formants 1 and 2. The concept of formants explains why the same pitch sounds different when played on similar woodwind instruments, like the clarinet and the saxophone — the instruments have different geometry, they amplify different frequencies, and result into different formants. It also explains why individuals of the same sex and gender do not sound the same when singing the same pitch (Pernet & Belin, 2012).

Voice gender perception. Before puberty, the voice of boys and girls sounds similar, with average f_o and formant frequencies significantly overlapping in these age ranges (Barreda & Assmann,

2021). Male puberty drives several biophysical changes, such as the vocal folds enlargement, the lowering of the larynx and subsequently the lengthening of the vocal tract, and as a result cisgender adult men have on average vocal tracts that are 15% larger than those of cisgender adult women (Fant, 1966). These sexually dimorphic anatomical features of the vocal tract drive the acoustic differences between adult male and female voices. However, it is still difficult to attest which acoustic parameter(s) is/are more relevant to the perception of gender in the human voice. A recent study by Neuhaus et al. (2024) took a step forward in that direction. They used computer-generated humanoid voice stimuli to investigate the dependence of two parameters proven to impact voice gender — i.e., the f_o and formant frequencies in the form of implied vocal tract length — and spectral tilt. They determined that increasing f_o and decreasing implied vocal tract length increase the likelihood of a speaker being perceived as female. In addition, an increased f_o and a steeper spectral tilt also increased the likelihood of a female-like judgment when the vocal tract length was relatively short. Thus, the use of computer-generated voice stimuli constitutes a unique opportunity to clarify which parameters (in isolation or combined) are critical to distinguish male from female voices. Furthermore, this clarification might be paramount to the preservation of self-expression during voice gender modification training.

Voice gender modification. To date, voice feminization programs have mostly been a complement to the surgical voice procedures, whose primary goal is to raise the speaking pitch. Surgical and non-surgical methods have been used to achieve voice feminization, namely directly raising f_o, altering formant frequency patterns, oral resonant therapy, semi-occluded vocal tract exercises, and 'diaphragmatic breathing' exercises (Schwarz et al., 2023). However, current literature lacks detailed descriptions of the voice programs conducted, particularly about the practices implemented. This pitfall contributes to the so-called reproducibility crisis (Camerer et al., 2018; Ioannidis, 2016). For instance, Hughes et al. (2021) conducted a retrospective study on the open quotient (i.e., a biophysical correlate of the spectral tilt) of transgender women after voice training. In this study, the voice training program was not described, and the patients did not all undergo the

same training (i.e., they differed in the program, instructor and/or institution), which is a common occurrence in the literature. In addition, few transgender voice studies collect voice gender heteroperception outcomes. Altogether, these issues decrease the future clinical usefulness of the current transgender voice studies.

This dissertation aims to: (1) address the voice features and voice experience of transgender men and women attending the Transgender Voice Consult of the Unit for Sex and Gender (*Unidade de Sexologia e Género)* at the *Unidade Local de Saúde Santo António* (USEG-ULSSA); (2) determine the acoustic predictors of voice gender heteroperception; and (3) inspect the voice effects, i.e., effects in voice features and voice experience, of the voice feminization training that is currently being conducted at the new Transgender Voice Consult of the USEG-ULSSA.

METHODS

Participants

This prospective study comprises four main groups: transgender women (TW), transgender men (TM), cisgender women (CW), and cisgender men (CM). In addition, an external listener group participated in the voice gender heteroperception task (for detailed description, please see "Voice gender heteroperception task").

Transgender participants. Transgender participants were Caucasians drawn from the patients of the Transgender Voice Consult at the USEG-ULSSA (July 2022 – December 2023; total number of patients: 44). The transgender sample comprises 17 TW (mean age = 28.5 ± 10.3 years, range = 19–50) and 10 TM (mean age = 24.9 ± 5.3 years, range = 19–35). Thirteen TW and 4 TM were excluded from the initial sample based on suspected atypical general cognition (*n* = 1), non-native speaking of European Portuguese (*n* = 4) and missing acoustic recording data (*n* = 12). None of the participants had received phonosurgery, and one TW had been subject to chondrolaryngoplasty. None of the participants were voice professionals or had voice pathology (see Procedures). Fourteen TW and eight TM were employed or students. The TW group had an average of 118 weeks

of GAHT at onboarding (M duration = 118.4 ± 152.3 weeks, range = .000–600.000; one TW participant initiated GAHT 1 week after onboarding) and the TM group had 101.10 weeks (M duration = 101.1 ± 89.7 weeks, range = 8.0–277.7).

Cisgender participants. Cisgender participants were Caucasians recruited through word-of-mouth. The sample comprises 21 CW (mean age = 29.7 ± 12.0 years, range = 19–59) and 17 CM (mean age = 28.5 ± 10.5 years, range = 19–51). Nineteen CW and seventeen CM were employed or students. See **Table I** for details.

The four groups did not differ significantly in age, education level, employment status or duration of GAHT. See **Table I** for detailed information.

Instruments

Self-reported voice gender perception and related quality of life were assessed at the onset of the study using three questionnaires: Voice Gender Modification Questionnaire, Voice Handicap Index (Jacobson et al., 1997; translated to European Portuguese by Guimarães & Abberton, 2004), and Trans Woman Voice Questionnaire (Dacakis et al., 2013; translated to and validated in European Portuguese by Alves et al., 2022).

Voice Gender Modification Questionnaire. Transgender participants were verbally asked: (1) the daily percentage of time that they perceive their voice as being congruent with their gender (0–100%) —hereafter, daily voice congruency (DVC)—; (2) whether they have voice gender modification habits (yes/no); and (3) whether they have physical and social conditions to practice their voice in private (yes/no).

Voice Handicap Index (VHI). The VHI is a widely used self-assessment instrument for measuring voice-related quality of life. It consists of 30 items rated on a four-point Likert scale and designed to capture concerns related to functional, physical, and emotional domains. The score ranges from 0 to 120. The higher the total score, the greater the negative impact of voice on the person's quality

of life. All participants completed the questionnaire during onboarding. Only the total score was used (min-max).

Trans Woman Voice Questionnaire (TWVQ). The TWVQ is an adaptation of the VHI to the transgender woman population. Like VHI, it has 30 items and has been translated and validated to European Portuguese. TW and CW completed the questionnaire as part of the onboarding procedure. Only the total score was used in this study.

Procedures

Transgender participants initiated the onboarding procedure at their first transgender voice consult at the USEG-ULSSA.

Transgender participants. Transgender participants were observed by the same otorhinolaryngologist. They completed a rigid fiberoptic videolaryngoscopy without stroboscopy (XION EndoSTROBE, type CD11F/R, XION GmbH Pankstraße, Berlin, Germany) to exclude observable voice pathology. TW and CW completed the VHI and TWVQ after their first otorhinolaryngology appointment at the transgender voice consult of the USEG-ULSSA. For all participants, voice samples were collected in a repurposed acoustically isolated audiometry booth in the otorhinolaryngology department using an Element Series EM-USB Condenser Microphone, a Power Studio pop filter, and the software Audacity (version 3.4.2.; Audacity Team, 2024) in mono mode with a sampling rate of 44.1 kHz and 16-bit quantization. Voice recordings were collected as part of the transgender voice consult protocol at USEG-ULSSA. From these, we selected the standard International Phonetic Alphabet (IPA) vowel /a/, the short phrase *Tigers are wild* ("Os tigres são selvagens"), and the phonetically balanced text "O Sol". The short phrase was selected from a validated set of Portuguese sentences developed for research on emotional prosody (Castro & Lima, 2010). Participants were instructed to speak with their day-to-day vocal disposition while speaking unhurriedly, in an intelligible manner, and at a comfortable voice amplitude, without consciously adding any emotional state to the sentence stimuli (i.e., emotionally neutral recording).

Voice recordings were manually trimmed in Audacity (version 3.4.2.; Audacity Team, 2024) and saved as high-quality uncompressed *wav* files; vowel recordings preserved voice onset and offset. TW participants repeated the onboarding procedure after discharge from training.

Cisgender participants. CW completed the TWVQ and VHI. CM completed the VHI. Both were submitted to the same voice recording session described above.

Voice acoustics

For voice stimuli, we extracted multiple acoustic measures using the MATLAB application VoiceSauce (Shue et al., 2011), namely average speaking fundamental frequency (f_o), the amplitude of the first harmonic corrected for formant frequencies (H1*), spectrum energy, the frequency of the third formant (F3) and the frequency of the fourth formant (F4). The Straight algorithm was used for f_o estimation, with settings of minimum f_o of 40 Hz and maximum f_o of 500 Hz. A window size of 25 ms, a 1 ms frame shift, and number of periods of harmonic estimation of 3 was used for harmonic extraction using the Straight algorithm. We selected the /a/ vowel recordings (over the /i/ vowel) based on its ubiquity in the European Portuguese dialect, and absence of lip rounding, and to prevent the narrowing of the vocal tract from distorting the relation between vocal tract length and formant frequencies. The phrase *Tigers are wild* ("Os tigres são selvagens") was selected because it was the only single-phrase stimulus that included an /a/ vowel. To prevent listener fatigue from lengthy stimuli, a fixed central segment of the text *The Sun* ("O Sol"; i.e., "Sem o brilhar do Sol, a Terra ficaria fria, sem plantas, mais pobre e menos bela") was used for further analysis. We extracted measures for every millisecond and computed median values of each parameter per stimulus.

Fundamental frequency. We extracted average f_o (Hz) from all stimuli. It is widely accepted that the human ear perceives pitch on a logarithmic scale. It is the basis behind scientific musical notation, music composition, and musical note interval perception. For this reason, the f_o values were converted from the linear Hertz scale to the logarithmic Musical Instrument Digital Scale (MIDI)

scale prior to analysis. This scale is often used in electronic music production. With this purpose, we used the following formula; the frequency of the musical note A4 (440 Hz) is the reference pitch for the base-two logarithmic function:

$$f_o\ (MIDI) = 12 \times \log_2 \frac{f_o\ (Hz)}{440\ Hz} + 69$$ (Equation 2. Conversion of f_o [Hz] to f_o [MIDI])

According to this approach, larger changes in Hertz are given smaller emphasis at higher pitches, and smaller changes in Hertz are given greater emphasis at higher pitches. In addition, the distance between any two adjacent MIDI values, for example, 52 (164.8 Hz) and 53 (174.6 Hz), is exactly one semitone, in accordance with scientific musical notation.

Spectral tilt. This parameter can be assessed using different types of indicators. The most common variant is H1-H2 (the difference in amplitude between the first and second harmonics). This measure is related to voice quality perception and lower values of H1-H2 are associated with lower glottal open quotient and male voices. A recent study by Chai & Garellek (2022) takes a deep dive into the history of spectral tilt measures and the rationale behind subtracting the amplitude of harmonic 2 (H2) from the amplitude of harmonic 1 (H1). In this study, they also propose a new parameter, residual H1, as a more theoretically sound and practical method of H1 normalization to sound pressure level. This method was verified by the authors through electroglottogram measurements, which is a well-established method for measuring glottal open quotient. A Source-Filter interaction is responsible for some changes in harmonic frequencies as they pass through the vocal tract. For this reason, harmonic frequencies can be corrected for the effect of formants. In the present study, we'll use the variable residual H1 corrected for formant frequencies and bandwidths (H1*).

Formants 3, 4 and implied vocal tract length (iVTL). Assuming a model of a cylindrical vocal tract that is open at one end (the lips) and closed at the opposite end (the glottis), the following formula can be used to estimate the vocal tract length, given a specific formant number *n* and frequency F_n and *c* equal to the speed of sound in the vocal tract (354 m/s):

$$iVTL = \frac{(2n-1)\times c}{2\times F_n}$$ (Equation 3. Implied vocal tract length)

Because the human vocal tract does not perfectly match this cylindrical model, we replaced the term *estimated* VTL by *implied* VTL, which mirrors the perceptual nature of this numerical value. The iVTL value used in the present study constitutes an arithmetic average of the iVTL predicted through F3 and that through F4.

Voice gender heteroperception task

We assessed voice gender perception using a task implemented via the online platform Qualtrics XM (Qualtrics XM®, Provo, UT, June 2024). Listeners were recruited through word-of-mouth, social media platforms, and associated hospital and university mailing lists. Exclusion criteria included age under 18 or over 100 years old, nationalities other than Portuguese, and low self-reported hearing acuity. Hearing acuity was self-reported on a 5-point Likert Scale (1 = poor; 5 = excellent). A score that was 2 standard deviations below the mean was considered as low hearing acuity. Listeners were not compensated for their participation.

Listeners completed an initial questionnaire that included anonymous ID, sex, age, country of origin, highest education level, current professional situation, hearing acuity, any neurologic or psychiatric disorders, and relevant medication.

The total set of voice stimuli (i.e., vowel, single-phrase, and phrase-in-text; please see "Procedures") included recordings from the 4 groups of participants. For the TW group, pre- and post-training recordings of the same type of voice stimulus were included. Two equidimensional groups of listeners were randomly established. Each group was randomly assigned a set of 108 voice stimuli from the total set of 216, ensuring that the selection was random and not sequential. The stimuli were presented in random order. Participants were instructed not to take breaks during the experiment. They rated each voice stimulus on a five-point Likert scale based on perceived sex of the speaker (-2 = *woman*; -1 = *probably a woman*; 0 = *androgynous / don't know*; 1 = *probably a man*; 2 = *man*). The terms *man* ("homem") and *woman* ("mulher") were used instead of *masculine*

("masculino") and *feminine* ("feminino") to avoid conflation with stereotypes of masculine or feminine speech. The mean voice gender heteroperception score for each voice stimulus was registered —hereafter, voice gender heteroperception. This variable will be used for testing associations with sociodemographic and acoustic variables.

A total of 175 participants completed the experiment; 11 were excluded due to nationalities other than Portuguese (n = 5), non-binary gender identity (n = 3), low hearing acuity (n = 2), and neurodegenerative disorder (n = 1). The final sample included 164 participants (113 women), mean age = 28.140 years ± 11.541, range = 18–62. A total of 101 had higher education, 44 secondary education, and 1 basic education. 57 were employed, 102 were students, and 5 were unemployed.

Voice training

A cisgender female in-hospital speech-language pathologist, who was external to the research team, selected 7 TW for flexible and individual voice training based on the clinical information provided. At pre-training, TW who were selected for voice training —hereafter, pre-training TW— did not differ from CW in age, t = .639, p = .528, BF_{10} = .451, education level, χ^2 = .000, p = 1.000, BF_{10} = .388, and employment status, χ^2 = .479, p = .787, BF_{10} = .342.

Training consisted of voice conditioning and voice feminization training. For voice conditioning, participants were instructed on basic anatomy and physiology of voice production and given health recommendations, including smoking cessation, conscious hydration, and not overdo the duration or frequency of voice production/training. Warning signs were also provided: pharyngeal globus or scratching sensation, pain, and vocal fatigue. Participants were advised to take vocal rest periods of up to an hour if they experienced vocal fatigue. For voice feminization training, participants were given instructions to: (1) speak at their most comfortable pitch in the speech typical laryngeal vibratory mechanism M1, (2) aim toward sympathetic vibrations localized to the maxilla, (3) facilitate better diction through greater amplitude of mouth movements, (4) modify their loudness (increasing or decreasing average loudness level and/or increasing loudness

flexibility), and (5) alter the melodic curve (e.g. increasing overall intonational variation while reducing downward intonational shifts; upper inflections at the ends of phrases; longer duration of phonemes; discursive pauses).

Voice conditioning exercises should be performed at least twice daily, with varying degrees of vocal tract semi-occlusion, ideally before and after prolonged voice use. Exercises included producing the IPA vowel /u/ and/or the consonant 'v' while, sequentially performing: (1) a constant pitch; (2) vocal sirens (an exercise where you raise and lower the pitch in a sinusoidal pattern); (3) successive arpeggios; (4) the song *Happy Birthday* in European Portuguese; (5) successive *crescendo-decrescendo* (an exercise where you raise and lower the loudness) without necessarily maintaining a constant pitch. The varying degrees of vocal tract semi-occlusion were defined by the speech therapist based on the progress and included successively: (1) phonating through a plastic straw submerged in water (at a ratio of 5-10 cm of submerged straw depth to 15 cm of water height); and (2) finger kazoo (phonating against an upright finger held against both lips). Voice feminization training included modeling and feedback (directly from the speech-language pathologist and/or via recorded voice samples), a progression of exercises from low-complexity to high-complexity speech, and the application of voice conditioning techniques during pitch and resonance modification exercises. Visual and auditory biofeedback via pitch and loudness analysis apps were used consistently.Based on clinical reasoning, participants progressed from these exercises to practicing automatic series (numbers, days of the week, and counting backwards) and colloquial phrases with increasing complexity (*good afternoon; good afternoon, therapist; good afternoon, therapist, how are you?; good afternoon, therapist, how is your day going*) with or without an oral semiocclusion mask.

During training sessions, the speech-language pathologist would continuously assess progress and adjust exercises (if necessary), or progress to the next series based on clinical reasoning. Conversation and role-playing were also incorporated into training sessions. Participants were encouraged to practice outside of training sessions with training partners (via WhatsApp).

Participants were instructed to monitor their pitch during the training session and every other day by sending in recordings of themselves reading, via email. This was analyzed and commented on as a continuous training feedback and adherence reinforcement. They were also instructed to monitor their loudness, while controlling distance from the microphone and background noise when recording. Participants who spoke in M2 were additionally instructed to avoid this approach by lowering their pitch and using the voice models (of family members or others) as reference.

Training was conducted until clinical discharge by the speech therapist. Clinical discharge occurred when clinical and personal meaningful outcomes had been reached (mean = 8.3 sessions; range = 6–12). The onboarding procedure was then repeated for participants who had undergone voice training.

Statistical analyses

Statistical analyses were conducted using the free, open-source software JASP (version 0.18.3; JASP Team, 2024). We conducted standard frequentist and Bayesian analyses. In addition to *p*-values, a Bayes Factor (BF) was estimated for each analysis using default priors; BFs consider the likelihood of the observed data under both the alternative and null hypotheses. BFs were interpreted according to Jeffreys' guidelines (Jarosz & Wiley, 2014; Figure 2). We present descriptive statistics for each variable, including mean (M), standard deviation (*SD*), and range. We used Pearson correlations and regression analyses to test associations between interest variables, and one-way ANOVAs and chi-square tests to compare groups. Paired sample t-tests were conducted to inspect for pre- to-post-training effects in voice gender heteroperception and acoustics.

RESULTS

Self-reported voice perception and voice-related quality of life

Voice Gender Modification Questionnaire. Fifteen TW and seven TM completed this questionnaire. TW and TM did not differ in their daily voice congruency, t = -1.266, p = .220; BF_{10} = .699. TW DVC was not correlated with GAHT duration, r = -.147, p = .601, BF_{10} = .361. There is moderate evidence towards a positive correlation between TM DVC and GAHT duration, r = .801, p = .030; BF_{10} = 3.241. A total of six TW (35.294%) and four TM (40.000%) reported voice feminization habits. Thirteen TW (76.471%) and eight TM (80.000%) reported being able to practice their voice in private.

VHI. TW presented an average score of 41.588 on VHI (SD = 27.900, range = 7–96), TM 47.000 (SD = 22, range = 22–91), CW 5.571 (SD = 4.697, range = 0–19), and CM 8.813 (SD = 12.156, range = 0–39). There were significant between-group differences in this variable, $F_{3,60}$ = 21.869, p < .001, η^2 = .522; BF_{10} > 100. TW presented higher VHI scores than CW, p < .001, BF_{10} > 100, and TM presented higher VHI scores than CM, p < .001, BF_{10} > 100. Scores were similar between CW and CM, p = .167, $BF_{10,U}$ = .523, and TW and TM, p = .116, $BF_{10,U}$ = .412. VHI scores were not correlated with GAHT duration for either TW, r = .110, p = .673, BF_{10} = .325 or TM, r = -.407, p = .243, BF_{10} = .710. One CM was excluded from this analysis due to lack of data.

TWVQ. TW presented an average score of 75.824 (SD = 21.475, range = 35–101) and CW 33.905 (SD = 4.170, range = 30–46), and the groups differed in this variable, t = 8.770, p < .001, d = 2.861; BF_{10} > 100. TWVQ scores did not correlate to GAHT duration, r = .003, p = .990, BF_{10} = .300.

Acoustic voice description

Relationships between the acoustic parameters for each voice stimuli and sociodemographic variables, duration of GAHT, DVC, VHI and TWVQ are described in **Table II;** for Group comparisons of acoustic parameters, please see **Table III.**

Mean speaking f_o. The average speaking f_o did not relate to any variable, except group. The groups presented significant differences in average speaking f_o for all stimulus types. For all stimulus types,

CW presented the highest average speaking f_o. Both TM and TW spoke at higher f_o than CM for all stimulus types. There were no significant differences in average speaking f_o between the TW and TM groups.

Residual H1.* Residual H1* presented significant associations with Group, DVC, VHI, and TWVQ. Regarding groups, they differed in residual H1* for all stimulus types (see **Tables II** and **III**). The largest difference in residual H1* was between the TW and CW groups for vowels and the CW and CM groups for single-phrase and phrase-in-text stimuli. There was no significant difference between CM and TM or TW for any stimulus type. TW and CW presented different residual H1* values for all stimulus types. DVC correlated with residual H1* of single-phrase stimuli in TM and VHI correlated with residual H1* in vowel and single-phrase stimuli for TW, but anecdotal evidence supported the association between VHI and residual H1* in vowel stimuli. TWVQ scores were correlated with vowel-stimulus residual H1* for TW.

Implied vocal tract length (iVTL). iVTL related to age, Group, DVC, VHI, and TWVQ. A positive correlation was found between age and iVTL for vowel stimuli. DVC correlated with iVTL for vowel and phrase-in-text stimuli in TW, while for TM and CW, an association was found between VHI and phrase-in-text iVTL. TWVQ scores were correlated with phrase-in-text iVTL for CW. Regarding Group comparisons, groups differed in iVTL for single-phrase and phrase-in-text stimuli. For single-phrase stimuli, CW presented lower iVTL than CM, TW, and TM. For phrase-in-text stimuli, CM presented the highest iVTL. We found no difference in iVTL between TW and TM for single phrases.

No association was found between the acoustic parameters and education level or employment status.

Predictors of voice gender heteroperception

Descriptive statistics and group comparisons for voice gender heteroperception are presented in **Table III**. TW had an average rating of *probably a man* for vowel stimuli, and *man for single-phrase*

and *phrase-in-text* stimuli. TM were consistently rated *probably a man*. For all stimuli, CW and CM had average ratings of *woman* and *man*, respectively. For all stimuli, the groups differed in voice gender perception (see **Table III**).

Relationships between voice gender heteroperception and sociodemographic variables, and voice-related questionnaires are described in **Table IV**. DVC was not related to voice gender heteroperception for vowel, single-phrase or phrase-in-text voice stimuli for TW or TM. Voice gender heteroperception correlated with VHI scores in all stimuli, but the pairwise multiple comparisons did not reach significance ($p > .05$ and/or $BF_{10} < 3$). TWVQ scores related to single-phrase stimuli for CW, but not for any other stimuli or group. Voice gender heteroperception was not associated with education level or employment status. For all voice stimuli, there was no relationship between voice gender heteroperception and duration of GAHT in TW or TM individuals. Please see **Tables V – VII** for relationships between voice gender heteroperception and average speaking f_o, residual H1*, and iVTL in the three types of stimuli**.** For vowel stimuli, voice gender heteroperception was correlated to average speaking f_o and residual H1*, but not iVTL. For single-phrase and phrase-in-text stimuli, voice gender heteroperception related to average speaking f_o, residual H1*, and iVTL. We conducted a regression analysis to inspect the variance in voice gender heteroperception that was predicted by sociodemographic and acoustic variables for the three voice stimuli. Results are presented in **Table VIII**. We found that sociodemographic variables and acoustic parameters explained 80.8%, 80.0%, and 80.4% of the variance in voice gender heteroperception in vowel, single-phrase, and phrase-in-text stimuli, respectively. Even controlling for age, education level, and employment status, average speaking f_o remained a significant predictor of voice gender heteroperception in vowel ($t = -14.982$, $p < .001$, $BF_{10} > 100$), single-phrase ($t = -10.163$, $p < .001$, $BF_{10} > 100$), and phrase-in-text stimuli ($t = -5.803$, $p < .001$, $BF_{10} > 100$), and iVTL for single-phrase ($t = 3.813$, $p < .001$, $BF_{10} = 46.504$) and phrase-in-text stimuli ($t = 3.919$, $p < .001$, $BF_{10} = 48.186$).

Pre- to post- training effects

At pre-training, TW who were selected for voice training —hereafter, pre-training TW— did not differ from CW in age, t = .639, p = .528, BF_{10} = .451, education level, χ^2 = .000, p = 1.000, BF_{10} = .388, and employment status, χ^2 = .479, p = .787, BF_{10} = .342. Pre-training TW and CW presented similar iVTL for vowel stimuli and residual H1* for single-phrase stimuli (see **Table IX** for further information).

Voice-related quality of life questionnaires. Neither VHI nor TWVQ scores improved as a result of training (VHI: t = .442, p = .674; BF_{10} = .383; TWVQ: t = 1.023, p = .346; BF_{10} = .530).

Voice gender heteroperception and acoustic parameters. In vowels, the voice of TW were assessed as more woman-like as a result of training for vowel stimuli, t = 3.472, p = .013, BF_{10} = 5.402, but not for single-phrase, t = 2.259, p = .065, BF_{10} = 1.644, or phrase-in-text stimuli, t = 1.628, p = .155, BF_{10} =.886. Despite the change in perceived voice gender for vowel stimuli in TW, their score remained different than that of CW, t = -15.932, p < .001, BF_{10} > 100. There were no significant pre- to post-training changes in acoustic parameters for vowel (average speaking f_o: t = -.923, p = .392, BF_{10} = 494; residual H1*: t = -.572, p = .588, BF_{10} = .372; iVTL: t = .491, p = .641, BF_{10} =.390) or single-phrase stimuli (average speaking f_o: t = -1.791, p = .124, BF_{10} = 1.304; residual H1*: t = -.302, p = .773, BF_{10} = .367, iVTL: t = .559, p = .597, BF_{10} = .402). Average speaking f_o increased for phrase-in-text stimuli, t = 1.628, p = .155, BF_{10} = .886, but remained lower than that of CW, t = -21.702, p < .001, BF_{10} > 100. Residual H1*, t = .353, p = .736, BF_{10} = .372, and iVTL, t = .559, p = .597, BF_{10} = .402, was similar pre- and post-training for phrase-in-text stimuli.

DISCUSSION

In the present study, we demonstrate that (1) TW and TM experience lower voice-related quality of life compared to their cisgender peers; (2) sociodemographic and acoustic variables accounted for over 80% of the variability in voice heteroperception, and f_o and iVLT were the main predictors; (3) after training, TW exhibited a more woman-like voice as perceived in the heteroperception of vowels, and a higher f_o in phrase-in-text stimuli, although TW voices remained different from those of CW; and (4) the training program was not useful in improving voice-related quality of life as assessed by the VHI and the TWVQ questionnaires. To our best knowledge, this is the first empirical research addressing voice gender modification in Portugal.

It is widely acknowledged that feminizing GAHT is not effective in modifying voice (Schwarz et al., 2023). Our results are in line with this assumption: TW and CM were rated similarly in the voice gender heteroperception task, and GAHT duration did not relate to voice gender heteroperception in TW. Conversely, masculinizing GAHT is a proven tool for voice masculinization, and this motivates a lower investment in non-hormonal voice masculinization interventions. Previous studies have reported that masculinizing GAHT is mostly effective at masculinizing f_o in TM (Hodges-Simeon et al., 2021; Ziegler et al., 2018). We found, however, that the f_o of TM was similar to that of CW. TM were consistently rated *probably female*, while CM were consistently rated *male*, and the perception of their voices (in the voice gender heteroperception task) was not correlated with GAHT duration for this group. This mismatch with the current literature might be a result of sample bias, because (1) reporting voice complaints was a criterion for referral to the USEG-ULSSA transgender voice consult and (2) most of our TM sample was above the threshold for voice masculinization (around 3 to 6 months) (Ziegler et al., 2018). Furthermore, our results point that TM present iVTL values halfway between those of CM and CW for single-phrase and phrase-in-text stimuli. These findings align well with those by Hodges-Simeon et al. (2021), who reported that the iVTL of their TM sample was longer than that of CW, but shorter than that of CM. Our findings thus

add to a growing body of literature that suggests that masculinizing GAHT may not always be sufficient for voice masculinization.

Furthermore, the results support the existence of a positive relationship between DVC and GAHT duration for TM but not TW. No relationship was found between DVC and gender heteroperception in TM or TW. The implication is that DVC may not result from the perception of voice gender by others, but instead from self-perception. This suggests that achieving voice gender congruence is a complex process influenced by factors beyond hormonal and voice changes, but also psychological factors related to gender dysphoria. Thus, personalized voice care is essential for transgender individuals, and may include cognitive behavioural therapy and additional approaches that allow to decrease voice dysphoria.

Regarding voice-related quality of life, TW and TM reported significantly higher VHI scores than the cisgender participants. This finding aligns well with previous research highlighting the negative impact of voice on the quality of life of transgender individuals (Hancock, 2017; Schwarz et al., 2023). Notably, there was no significant difference in VHI scores between TW and TM, suggesting that voice-related concerns are prevalent across the transgender spectrum, regardless of gender identity. Yet, the hypothesis that this result derives from a sampling bias cannot be discarded. Nonetheless, this points towards the relevance of patient-centred voice care in transgender healthcare, that addresses the specific challenges faced by both TW and TM in achieving voice gender congruence and voice comfort. This could, therefore, improve voice-related quality of life.

When the topic is voice gender perception, for vowels, perception was mainly predicted by average speaking f_o, while for single-phrase and phrase-in-text stimuli, both average speaking f_o and iVTL were significant predictors. This result is in line with previous research (Neuhaus et al., 2024) and suggests that, while pitch is a primary cue for gender perception, other acoustic features, such as vocal tract length, may also play a key role, especially in more ecological speech fragments. Our study does not clarify why voice gender is more easily perceived in longer stimuli. Two main

hypotheses have been proposed: (1) longer stimuli allow for more linguistic and acoustic content to be perceived, or (2) suprasegmental features contribute to gender perception. Clarifying on this matter has implications for voice feminization training, as it suggests that interventions should incorporate phonation of encompassing complexity to achieve optimal results for different phonatory contexts.

Unlike prior research that reported positive effects of voice training on voice-related quality of life and perceived voice gender in TW (Schwarz et al., 2023), our results suggest otherwise, i.e., TW presented post-training voice features different from CW and no benefits were found on voice-related quality of life. This discrepancy in results may be related to specific training methods, for instance, the weight of voice conditioning vs. voice training in the training protocol. Duration and intensity of the training, individual differences in response to training should also be considered. Further research, that includes a control group and a larger sample size, is mandatory to identify the most effective voice training approaches for TW and to better understand which factors influence individual responses to training.

As we previously mentioned, the recent work by Neuhaus et al. (2024), and that used computer-generated humanoid voice stimuli, determined that increasing f_o and decreasing implied vocal tract length increased the likelihood of a speaker being perceived as female. In addition, an increased f_o and a steeper spectral tilt also increased the likelihood of a female-like judgment when the vocal tract length was relatively short. Further research is required to attest on this matter with natural human voices, and to test models that consider diverse acoustic parameters in parallel, as well as their interplay (e.g., using mediation models).

This study has several limitations. The sample size was small to test the effectiveness of the voice feminization training program, and participants were predominantly Caucasian, which may limit the generalization of the findings. Additionally, the study did not include an active control group for the voice feminization training program, making it difficult to isolate the effects of the training from other factors (e.g., motivation). Formant measurement also introduces a degree of

error to two variables in this study: H1*, and iVTL. Through spectrographic analysis, we identified that formants were not always correctly tracked algorithmically, and this may be partly related to the different pronunciations of the vowel /a/ in the European Portuguese language and dialect. The introduction of nasality to this phoneme may create an antiformant, causing, for example, F4 to split into two frequency peaks, confounding automatic formant frequency extraction algorithms. This study also did not control the impact of listener's gender on their voice gender perception. This is especially relevant considering the proportion of female listeners in our heteroperception experiment. Lastly, the most ecological voice stimuli used in the present study is the text "The Sun", but free speech or dialogue should be the target for studies inspecting voice features, perception, and modification.

CONCLUSION

The findings of this study have important implications for transgender voice care and research. They emphasize the need for a comprehensive approach to voice modification that considers diverse acoustic parameters, and not only f_o. These findings also highlight the limitations of GAHT in achieving desired voice outcomes for TW, but also for TM, and suggest that additional interventions, such as voice training programs, are required. The limited efficacy of the current voice feminization training program unveils that further research is necessary in order to develop more effective interventions that address multiple acoustic parameters and improve voice-related quality of life.

Ethical approval

This study was approved by the executive board of the ULSSA, the ULSSA/ICBAS Ethics Committee, and the Centro Académico Clínico ICBAS-CHP. Written informed consent was obtained from all participants prior to data collection, in accordance with the General Data Protection Regulation and the 1964 Helsinki declaration and its later amendments or comparable ethical standards; for the external listener group informed consent was collected online, before data collection.

Disclosure statement: The authors state there are no competing interests to disclose.

REFERENCES

Alves, D. C., Correia, P., & Moreira, K. (2022). *Trans Woman Voice Questionnaire: authorised European Portuguese translation [versão traduzida para português europeu de Dacakis, G., & Davies, S. (2012). Transsexual Voice Questionnaire for male to female (TVQ-MtF) Melbourne, Australia: La Trobe University]*. http://hdl.handle.net/10400.26/45104

Barboza, G. E., Dominguez, S., & Chance, E. (2016). Physical victimization, gender identity and suicide risk among transgender men and women. *Preventive Medicine Reports, 4*, 385–390. https://doi.org/10.1016/j.pmedr.2016.08.003

Barreda, S., & Assmann, P. F. (2021). Perception of gender in children's voices. *The Journal of the Acoustical Society of America*, *150*(5), 3949–3963. https://doi.org/10.1121/10.0006785

Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., den Heijer, M., & T'Sjoen, G. (2017). Self-perception of voice in transgender persons during cross-sex hormone therapy. *Laryngoscope*, *127*(12), 2796–2804. https://doi.org/10.1002/lary.26716

Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods, 42*(1), 74–81. https://doi.org/10.3758/BRM.42.1.74

Chai, Y., & Garellek, M. (2022). On H1–H2 as an acoustic measure of linguistic phonation type. *The Journal of the Acoustical Society of America*, *152*(3), 1856–1870. https://doi.org/10.1121/10.0014175

da Silva, I. C. B., de Araújo, E. C., Santana, A. D. da S., Moura, J. W. da S., Ramalho, M. N. de A., & de Abreu, P. D. (2022). Gender violence perpetrated against trans women. In *Revista Brasileira de Enfermagem* (Vol. 75). Associacao Brasilerira de Enfermagem. https://doi.org/10.1590/0034-7167-2021-0173

Dacakis, G., Davies, S., Oates, J. M., Douglas, J. M., & Johnston, J. R. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *Journal of Voice*, *27*(3), 312–320. https://doi.org/10.1016/j.jvoice.2012.11.005

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *STL-QPSR*, *4*, 22–30. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=67951c34f793c4311f9cc2cc016aad57c2ac8cff

Guimarães, I., & Abberton, E. (2004). An investigation of the Voice Handicap Index with Speakers of Portuguese: Preliminary data. *Journal of Voice*, *18*(1), 71–82. https://doi.org/10.1016/j.jvoice.2003.07.002

Hancock, A. B. (2017). An ICF Perspective on Voice-related Quality of Life of American Transgender Women. *Journal of Voice*, *31*(1), 115.e1-115.e8. https://doi.org/10.1016/j.jvoice.2016.03.013

Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Groll, M. D., Stepp, C. E., Carré, J. M., & Arnocky, S. A. (2021). Testosterone therapy masculinizes speech and gender presentation in transgender men. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-82134-2

Hughes, C. K., McGarey, P., Morrison, D., Gawlik, A. E., Dominguez, L., & Dion, G. R. (2021). Vocal Fold Thinning in Transgender Patients. *Journal of Voice*, *37*(6), 957–962. https://doi.org/10.1016/j.jvoice.2021.06.026

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and Validation. *American Journal of Speech-Language Pathology*, *6*(3), 66–70. https://doi.org/10.1044/1058-0360.0603.66

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *Journal of Problem Solving*, *7*(1), 2–9. https://doi.org/10.7771/1932-6246.1167

Kim, H. T. (2020). Vocal feminization for transgender women: Current strategies and patient perspectives. In *International Journal of General Medicine* (Vol. 13, pp. 43–52). Dove Medical Press Ltd. https://doi.org/10.2147/IJGM.S205102

Mészáros, K., Csokonai Vitéz, L., Szabolcs, I., Góth, M., Kovács, L., Görömbei, Z., & Hacki, T. (2005). Efficacy of Conservative Voice Treatment in Male-to-Female Transsexuals. *Folia Phoniatrica et Logopaedica*, *57*(2), 111–118. https://doi.org/10.1159/000083572

Neuhaus, T. J., Scherer, R. C., & Whitfield, J. A. (2024). Gender Perception of Speech: Dependence on Fundamental Frequency, Implied Vocal Tract Length, and Source Spectral Tilt. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2024.01.014

Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization. *Frontiers in Psychology*, *3*(FEB). https://doi.org/10.3389/fpsyg.2012.00023

Qualtrics XM (2024).

Schwarz, K., Cielo, C. A., Spritzer, P. M., Villas-Boas, A. P., Costa, A. B., Fontanari, A. M. V., Costa Gomes, B., da Silva, D. C., Schneider, M. A., & Lobato, M. I. R. (2023). A speech therapy for transgender women: an updated systematic review and meta-analysis. In *Systematic Reviews* (Vol. 12, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13643-023-02267-5

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A Program for Voice Analysis. In *ICPhS XVII Regular Session Hong Kong*.

Simpson, A. P. (2009). Phonetic differences between male and female speech. *Linguistics and Language Compass*, *3*(2), 621–640. https://doi.org/10.1111/j.1749-818X.2009.00125.x

Team, A. (2024). *Audacity® [Computer software]*. https://audacity.sourceforge.net/

Team, J. (2024). *JASP [Computer software]*. https://jasp-stats.org/

Titze, I. (2009). *How Are Harmonics Produced at the Voice Source?* (pp. 575–576). https://vocology.utah.edu/_resources/documents/how_are_harmonics_produced_titze.pdf

Van Borsel G. De Cuypere, R. R. J. (2000). Voice problems in female-to-male transsexuals. *International Journal of Language & Communication Disorders*, *35*(3), 427–442. https://doi.org/10.1080/136828200410672

Ziegler, A., Henke, T., Wiedrick, J., & Helou, L. B. (2018). Effectiveness of testosterone therapy for masculinizing voice in transgender patients: A meta-analytic review. In *International Journal of Transgenderism* (Vol. 19, Issue 1, pp. 25–45). Routledge. https://doi.org/10.1080/15532739.2017.1411857

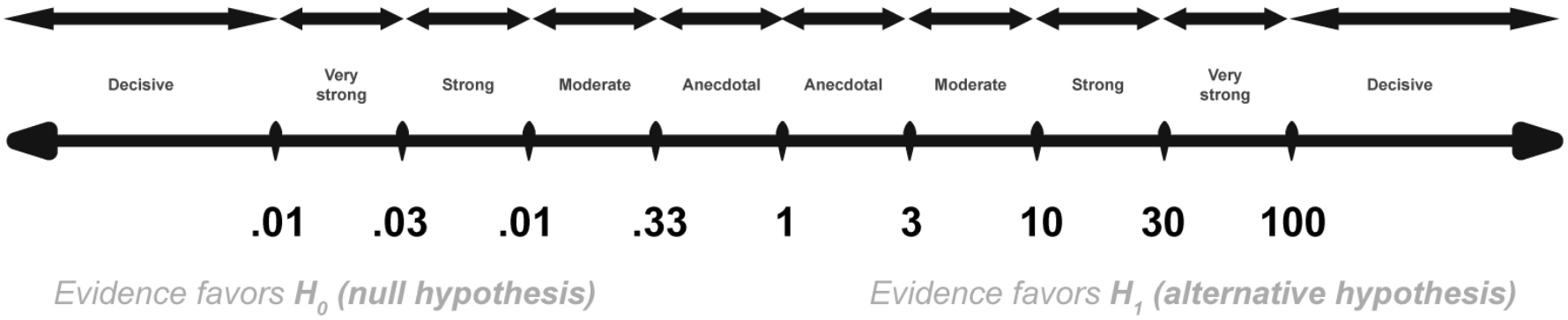


FIGURE 1. Interpretation of BF_{10} values according to Jeffreys' guidelines (Jarosz & Wiley, 2014).

Table I. Sociodemographic characteristics and group comparison of transgender women (TW), transgender men (TM), cisgender women (CW), and cisgender men (CM).

	TW (*n* = 17)	**TM (*n* = 10)**	**CW (*n* = 21)**	**CM (*n* = 17)**	**test, *p*, BF_{10}**
Age (years)	28.471 ± 10.308	24.900 ± 5.301	29.667 ± 11.964	28.529 ± 10.483	.483, .695 *.135*
Education level (basic/secondary/higher)	1/10/6	2/8/0	0/9/12	1/7/12	12.854, .045 *1.382*
Employment status (employed/student/unemployed)	9/5/3	5/3/2	10/9/2	9/8/0	4.453, .616 *0.034*
Duration of GAHT (weeks)	118.378 ± 152.314	101.100 ± 89.698	N/A	N/A	.106, .748 *.381*

TW. Transgender women. *TM.* Transgender men. *CW.* Cisgender women. *CM.* Cisgender men. *GAHT.* Gender-affirming hormone therapy. *N/A.* Not applicable. We present Pearson correlations for age and duration of GAHT, and chi-square tests for education level and employment status. BF_{10} values are indicated in italics.

Table II. Relationships between age, group, education level, employment status, duration of gender-affirming hormone therapy (GAHT), Daily Voice Congruency (DVC), Voice Handicap Index (VHI), Trans Woman Voice Questionnaire (TWVQ), and average speaking f_o, residual H1* and iVTL in vowel, single-phrase, and phrase-in-text stimuli.

	Vowel stimuli			**Single-phrase stimuli**			**Phrase-in-text stimuli**		
	Average speaking f_o	**Residual H1***	**iVTL**	**Average speaking f_o**	**Residual H1***	**iVTL**	**Average speaking f_o**	**Residual H1***	**iVTL**
Age	-.065 *.176*	-.137 *.155*	.311* *3.455*	-.096 *.205*	.089 *.198*	-.054 *.170*	-.045 *.165*	.150 *.311*	-.023 *.157*
Group	63.072*** *> 100*	4.311*** *6.722*	.814 *.195*	43.641*** *> 100*	8.045*** *> 100*	11.689*** *> 100*	47.799*** *> 100*	11.022*** *> 100*	41.252*** *> 100*
Education level	.827 *.255*	.427 *.223*	.108 *.178*	.370 *.203*	.012 *.170*	1.017 *.291*	.426 *.213*	.009 *.169*	.090 *.176*
Employment status	.690 *.252*	.514 *.227*	.039 *.156*	.544 *.218*	.832 *.283*	.188 *.168*	.645 *.236*	.022 *.153*	.978 *.167*
Duration of GAHT	TW: -.228 *.429* TM: .122, *.407*	TW: .148, *.306* TM: -.069, *.393*	TW: .230, *.432* TM: .070, *.393*	TW: -.198, *.392* TM: -.046, *.389*	TW: .183, *.377* TM: .427, *.760*	TW: .346, *.707* TM: .297, *.528*	TW: -.016, *.300* TM: -.119, *.406*	TW: .497*, *2.019* TM: .101, *.400*	TW: .262, *.484* TM: .124, *.407*
DVC	TW: .409, *.912* TM: .007, *.457* TW+TM: .222, *.421*	TW: -.370, *.743* TM: .427, *.681* TW+TM: -.204, *.390*	TW: -.764**, *46.983* TM: -.014, *.457* TW+TM: -.536*, *5.900*	TW: .166, *.374* TM: -.227, *.508* TW+TM: .042, *.269*	TW: .153, *.365* TM: .827*, *4.081* TW+TM: .239, *.454*	TW: -.477, *1.399* TM: .488, *.783* TW+TM: -.058, *.273*	TW: .068, *.327* TM: -.245, *.517* TW+TM: -.030, *.266*	TW: -.131, .352 TM: .243, .516 TW+TM: -.119, .301	TW: -.573*, *3.116* TM: .601, *1.101* TW+TM: -.159, *.334*
VHI	TW: -.076, *.311* TM: .051, *.390*	TW: .516*, *2.387* TM: -.016, *.387*	TW: .235, *.440* TM: .123, *.407*	TW: -.106, *.323* TM: .352, *.603*	TW: -.549*, *3.326* TM: -.448, *.820*	TW: -.254, *.470* TM: -.414, *.726*	TW: -.110, *.325* TM: .372, *.639*	TW: -.243, *.452* TM: .389, *.671*	TW: .198, *.392* TM: -.734*, *4.984*

	CW: .018, *.278* CM: .381, *.785* TW+TM+CW+C M: -.158, *.331*	CW: .414, *1.292* CM: -.060, *.325* TW+TM+CW+C M: .204, *.543*	CW: .297, *.590* CM: -.115, *.344* TW+TM+CW+C M: -.098, *.210*	CW: -.016, *.277* CM: .201, *.403* TW+TM+CW+C M: -.169, *.368*	CW: -.140, *.325* CM: -.305, *.557* TW+TM+CW+C M: -.014, *.159*	CW: -.342, *.767* CM: .268, *.489* TW+TM+CW+C M: .127, *.254*	CW: .137, *.323* CM: .103, *.339* TW+TM+CW+C M: -.161, *.339*	CW: -.080, *.292* CM: -.156, *.367* TW+TM+CW+C M: .149, *.304*	CW: .538*, *4.536* CM: -.371, *.746* TW+TM+CW+C M: .054, *.173*
TWVQ	TW: -.150, *.349* CW: .143, *.328*	TW: .550*, *3.363* CW: .436, *1.555*	TW: .411, *1.040* CW: .094, *.298*	TW: -.070, *.310* CW: .020, *.278*	TW: -.379, *.851* CW: -.067, *.287*	TW: -.051, *.305* CW: .314, *.647*	TW: -.106, *.323* CW: .149, *.332*	TW: -.093, *.318* CW: -.260, *.491*	TW: .303, *.573* CW: .667**, *34.115*

DVC. Daily voice congruency. f_o. Fundamental frequency. *GAHT*. Gender-affirming hormone therapy. *H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. *iVTL*. Implied vocal tract length. *TW*. Transgender women. *TM*. Transgender men. *TWVQ*. Trans Woman Voice Questionnaire. *VHI*. Voice Handicap Index. For age, duration of GAHT, VHI, and TWVQ values represent Pearson correlations. For group, education level and employment status, values represent ANOVA *F*-values. BF_{10} values are indicated in italics.

* $p < .05$; ** $p < .01$*** $p < .001$

Table III. Descriptive statistics and group comparison of voice gender heteroperception and acoustic parameters of vowel, single-phrase, and phrase-in-text stimuli in transgender women (TW), transgender men (TM), cisgender women (CW), and cisgender men (CM).

	TW ***n* = 17**	**TM** ***n* = 10**	**CW** ***n* = 21**	**CM** ***n* = 17**	**F, BF_{10} or Comparison, $BF_{10,U}$**
Vowel stimuli					
Voice gender heteroperception	.960 ± .806	.848 ± 1.062	-1.544 ± .323	1.842 ± .087	108.983***, *>100*
	(-.905 – 1.778)	(-1.108 – 1.797)	(-1.911 – -.595)	(1.611 – 1.946)	TW vs TM, *.157*
					*TW vs CW****, *> 100*
					TW vs CM***, *99.618*
					TM vs CW**, *> 100*
					TM vs CM***, *17.649*
					CW vs CM***, *> 100*
Average speaking f_o	50.148 ± 3.119	49.194 ± 2.645	55.920 ± 1.787	45.558 ± 1.868	63.072***, *> 100*
	(41.642 – 56.033)	(44.236 – 53.775)	(52.158 – 58.759)	(42.795 – 48.856)	TW vs TM, *.193*
					TW vs CW***, *> 100*
					TW vs CM, *> 100*
					TM vs CW***, *> 100*
					TM vs CM`***, *32.776*
					CW vs CM***, *> 100*

Residual H1*	26.638 ± 3.395	23.761 ± 4.552	23.113 ± 3.684	26.542 ± 3.430	4.311**, *6.722*
	(21.249 – 32.236)	(15.129 – 17.740)	(15.050 – 29.020)	(19.201 – 31.722)	TW vs TM, *.1.277*
					TW vs CW**, *9.394*
					TW vs CM, *.330*
					TM vs CW, *.382*
					TM vs CM, *1.166*
					CW vs CM**, *7.715*
iVTL	16.748 ± .816	16.282 ± .748	17.617 ± 3.943	16.912 ± 1.391	*.814, .195*
	(15.440 – 18.173)	(15.129 – 17.740)	(14.709 – 27.406)	(13.678 – 19.377)	TW vs TM, *.804*
					TW vs CW, *.432*
					TW vs CM, *.352*
					TM vs CW, *.538*
					TM vs CM, *.688*
					CW vs CM, *.384*
Single-phrase stimuli					
Voice gender heteroperception	1.565 ± .610	1.045 ± .958	-1.790 ± .250	1.935 ± .040	219.739***, *> 100*
	(-.611 – 1.946)	(-.770 – 1.867)	(-1.986 – -.978)	(1.838 – 1.973)	TW vs TM*, *1.071*
					TW vs CW***, *> 100*
					TW vs CM*, *3.255*
					TM vs CW***, *> 100*
					TM vs CM***, *41.583*

					CW vs CM***, *> 100*
Average speaking f_o	48.283 ± 2.519	49.235 ± 3.276	54.676 ± 2.291	45.708 ± 2.262	*43.641****, *> 100*
	(44.015 – 52.952)	(44.850 – 54.617)	(49.497 – 58.132)	(40.903 – 48.931)	TW vs TM, *.478*
					TW vs CW***, *> 100*
					TW vs CM**, *10.959*
					TM vs CW***, *> 100*
					TM vs CM***, *13.356*
					CW vs CM***, *> 100*
Residual H1*	26.005 ± 3.266	25.074 ± 2.879	21.889 ± 3.536	26.548 ± 3.033	*8.045****, *> 100*
	(16.714 – 29.695)	(17.756 – 27.259)	(11.937 – 27.058)	(18.623 – 29.651)	TW vs TM, *.450*
					TW vs CW***, *39.949*
					TW vs CM, *.362*
					TM vs CW*, *3.123*
					TM vs CM, *.642*
					CW vs CM***, *177.412*
iVTL	16.084 ± .737	16.140 ± .892	15.133 ± .500	16.330 ± .701	11.689***, *> 100*
	(14.778 – 17.281)	(15.163 – 17.955)	(14.316 – 15.915)	(14.598 – 17.434)	TW vs TM, .371
					TW vs CW***, *> 100*
					TW vs CM, *.482*
					TM vs CW***, *29.510*
					TM vs CM, *.422*

					CW vs CM***, *> 100*
Phrase-in-text stimuli					
Voice gender heteroperception	1.529 ± .611	1.092 ± 1.214	-1.787 ± .302	1.935 ± .043	160.429***, *> 100*
	(-.122 – 1.944)	(-1.700 – 1.959)	(-2.000 – -.770)	(1.822 – 2.000)	TW vs TM, *.648*
					TW vs CW***, *> 100*
					TW vs CM, *2.052*
					TM vs CW***, *> 100*
					TM vs CM***, *6.234*
					CW vs CM***, *> 100*
Average speaking f_o	48.242 ± 2.457	48.958 ± 3.124	54.452 ± 2.157	45.359 ± 2.207	*47.799****, *> 100*
	(43.970 – 52.749)	(44.617 – 54.642)	(49.865 – 58.889)	(40.431 – 48.901)	TW vs TM, *.431*
					TW vs CW***, *> 100*
					TW vs CM***, *29.529*
					TM vs CW***, *> 100*
					TM vs CM***, *8.080*
					CW vs CM***, *> 100*
Residual H1*	26.300 ± 1.664	25.112 ± 1.425	23.232 ± 1.893	26.545 ± 2.630	11.022***, *> 100*
	(23.776 – 28.621)	(22.268 – 26.829)	(19.724 – 27.405)	(18.035 – 28.950)	TW vs TM, *1.294*
					TW vs CW***, *> 100*
					TW vs CM, *.342*
					TM vs CW*, *5.252*

					TM vs CM, *.903*
					CW vs CM***, *> 100*
iVTL	16.312 ± .439	15.968 ± .618	15.287 ± .433	16.882 ± .366	41.252***, *> 100*
	(15.231 – 17.121)	(15.089 – 16.856)	(14.529 – 16.581)	(16.303 – 17.533)	TW vs TM, *1.016*
					TW vs CW***, *> 100*
					TW vs CM***, *96.819*
					TM vs CW***, *24.551*
					TM vs CM***, *> 100*
					CW vs CM***, *> 100*

CM. Cisgender men. *CW*. Cisgender women. f_o. Fundamental frequency. *iVTL*. Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. *TM*. Transgender men. *TW*. Transgender women. For selected TW before *vs* after training, paired samples *t*-tests were performed. BF_{10} values are indicated in italics.

Table IV. Relationships between voice gender heteroperception in vowel, single-phrase, and phrase-in-text stimuli and sociodemographic variables.

	Vowel voice gender heteroperception	**Single-phrase voice gender heteroperception**	**Phrase-in-text voice gender heteroperception**
Age	.033 *0.160*	-.035 *0.161*	-.041 *0.163*
Group	108.983*** *> 100*	219.739*** *> 100*	160.429*** *> 100*
Education level	.702 *.252*	.712 *.263*	.623 *.262*
Employment status	.549 *.227*	.160 *.171*	.240 *.180*
Duration of GAHT	TW: .039, *.303* TM: .312, *.546*	TW: .183, *.377* TM: .117, *.405*	TW: .187, *.381* TM: *.228*, *.463*
DVC	TW: -.516*, *1.880* TM: .437, *.696*	TW: -.178, *.383* TM: .211, *.500*	TW: -.371, *.748* TM: .307, *.556*

VHI	TW: .420, *1.106* TM: -.133, *.410* CW: .181, *.364* CM: -.270, *.492* TW+TM+CW+CM: .322*, *3.834*	TW: .224, *.423* TM: -.140, *.413* CW: .193, *.378* CM: -.157, *.368* TW+TM+CW+CM: .371**, *11.646*	TW: .313, *.598* TM: -.239, *.472* CW: .170, *.352* CM: .116, *.344* TW+TM+CW+CM: .364**, *9.937*
TWVQ	TW: .412, *1.046* CW: .067, *.287*	TW: .041, *.303* CW: .448*, *1.736*	TW: .249, *.462* CW: .314, *.649*

GAHT. Gender-affirming hormone therapy. BF_{10} values are indicated in italics. We conducted Pearson correlations for age and duration of GAHT, and one-way ANOVAs for group, education level and employment status.
*** $p < .001$.

Table V. Relationships between voice gender heteroperception and average speaking f_o, residual H1*, and implied vocal tract length (iVTL) for vowels.

STIMULUS: Vowel	**1**	**2**	**3**	**4**
1. Voice gender heteroperception		-.904** *> 100*	.346** *7.672*	-.071 *0.180*
2. Average speaking f_o			-.316* *3.886*	.054 *.169*
3. Residual H1*				.005 *.155*
4. iVTL				

f_o. Fundamental frequency. *GAHT*. Gender-affirming hormone therapy. *iVTL*. Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. BF_{10} values are indicated in italics.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table VI. Relationships between voice gender heteroperception and average speaking f_o, residual H1*, and implied vocal tract length (iVTL) for single-phrase stimuli.

STIMULUS: Single-phrase	**1**	**2**	**3**	**4**
1. Voice gender heteroperception		-.871*** *> 100*	.544*** *> 100*	.634*** *> 100*
2. Average speaking f_o			-.630*** *> 100*	-.508*** *> 100*
3. Residual H1*				.429*** *78.743*
4. iVTL				

f_o. Fundamental frequency. *GAHT*. Gender-affirming hormone therapy. *iVTL*. Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. BF_{10} values are indicated in italics.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table VII. Relationships between voice gender heteroperception and average speaking f_o, residual H1*, and implied vocal tract length (iVTL) for phrase-in-text stimulus.

STIMULUS: Phrase-in-text	**1**	**2**	**3**	**4**
1. Voice gender heteroperception		-.875*** *> 100*	.576*** *> 100*	.810*** *> 100*
2. Average speaking f_o			-.625*** *> 100*	-.763*** *> 100*
3. Residual H1*				.472*** *> 100*
4. iVTL				

f_o. Fundamental frequency. *GAHT*. Gender-affirming hormone therapy. *iVTL*. Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. BF_{10} values are indicated in italics.
* $p < .05$; ** $p < .01$; *** $p < .001$.

Table VIII. Summary of the linear regression for voice gender heteroperception in vowel, single-phrase, and phrase-in-text stimuli considering sociodemographic and acoustic parameters as predictors.

	Vowel stimuli			**Single-phrase stimuli**			**Phrase-in-text stimuli**		
	B	***SE B***	***β***	***B***	***SE B***	***β***	***B***	***SE B***	***β***
Age	.002	.010	.011	-.014	.011	-.082	-.008	.011	-.051
Education level	.170	.154	.070	-.028	.148	-.011	.150	.174	.054
Employment status	-.048	.128	-.022	.074	.175	.027	-.057	.149	-.023
Average speaking f_o	-.285	.019	-.882***	-.304	.030	-.783***	-.228	.039	-.576***
Residual H1*	.027	.022	.071	-.025	.033	-.055	.044	.051	.062
iVTL	-.019	.037	-.031	.517	.136	.257***	.763	.195	.342***
R^2	.826			.819			.823		
Adjusted R^2	.808			.800			.804		
***F*-change in R^2**	45.943***			43.766***			44.806***		
BF_{10}	> 100			> 100			> 100		

f_o. Fundamental frequency. *iVTL.* Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies.
*** $p < 0.001$.

Table IX. Descriptive statistics and group comparison of voice gender heteroperception and acoustic parameters of vowel, single-phrase, and phrase-in-text stimuli in cisgender women (CW) and pre- and post-training transgender women (TW pre and TW post).

	CW (*n* = 21)	**TW pre-training (*n* = 7)**	**TW post-training (*n* = 7)**	**CW vs TW pre**	***TW pre vs TW post***	***TW post vs CW***
Vowel stimuli						
Mean voice gender heteroperception scores	-1.544 ± 0.323	1.312 ± .367	0.930 ± 0.447	-19.609, <.001	3.472, .013	-15.932, < .001
				> 100	*5.402*	*> 100*
Average speaking f_o	55.920 ± 1.787	50.093 ± 1.338	50.744 ± 2.252	7.880, < .001	-.923, .392	6.228, < .001
				> 100	*.494*	*> 100*
Residual H1*	23.113 ± 3.684	27.166 ± 3.047	28.062 ± 3.003	-2.618, .015	-.572, .588	-3.205, .004
				3.834	*.372*	*10.792*
iVTL	17.616 ± .943	17.037 ± .541	16.916 ± .431	.382, .705	.491, .641	.462, .648
				.412	*.390*	*.422*
Single-phrase stimuli						
Mean voice gender heteroperception scores	-1.790 ± .250	1.737 ± .176	1.216 ± .597	-34.387, < .001	2.259, .065	-19.074, < .001
				> 100	*1.644*	*> 100*
Average speaking f_o	54.676 ± 2.291	47.907 ± 1.693	49.189 ± 2.652	7.155, < .001	-1.791, .124	5.284, < .001
				> 100	*1.034*	*> 100*

Residual H1*	21.889 ± 3.536	25.049 ± 4.702	25.573 ± 2.241	-1.887, .070 *1.326*	-.302, .773 *.367*	-2.571, .016 *3.553*
iVTL	15.133 ± .500	16.053 ± .850	16.204 ± .244	-3.520, .002 *19.886*	.559, .597 *.402*	-5.410, < .001 *> 100*
Phrase-in-text stimuli						
Mean voice gender heteroperception scores	-1.787 ± .302	1.762 ± .172	1.436 ± .445	-29.321, < .001 *> 100*	1.628, 0.155 *0.886*	-21.702, < .001 *> 100*
Average speaking f_o	52.452 ± 2.157	47.487 ± 1.815	50.025 ± 2.732	7.661, < .001 *> 100*	-3.044, .023 *3.588*	4.405, < .001 *> 100*
Residual H1*	23.232 ± 1.893	25.977 ± 1.820	25.659 ± 1.803	-3.352, .002 *14.291*	.353, .736 *.372*	-2.969, .006 *6.998*
iVTL	15.287 ± .433	1.762 ± .172	16.068 ± .242	-6.465, < .001 *> 100*	2.454, .050 *1.999*	-4.507, < .001 *> 100*
VHI	5.571 ± 4.697	51.714 ± 32.217	47.857 ± 20.868	-6.602, < .001 *> 100*	.442, .674 *.383*	-8.940, < .001 *> 100*
TWVQ	33.905 ± 4.170	84.714 ± 18.918	80.286 ± 17.095	-11.884, < .001 *> 100*	1.023, .346 *.530*	-11.821, <.001 *> 100*

CM. Cisgender men. *CW*. Cisgender women. f_o. Fundamental frequency. *iVTL*. Implied vocal tract length. *Residual H1**. Residual of the amplitude of the first harmonic corrected for formant frequencies. *TM*. Transgender men. *TW*. Transgender women. For selected TW pre- and post-voice training *vs* CW, values represent independent samples *t*-tests results. For selected TW pre- *vs* post training, paired samples *t*-tests were performed. BF_{10} values are indicated in italics.

INSTITUTO DE **CIÊNCIAS BIOMÉDICAS ABEL SALAZAR**